

Module 5

Data Source : Access table

Software : Pentaho Data Integration Software and Oracle Database.

Requirements : First we need to load access file and the check the filters and add validations.

1. Step 1 : Filter Rows according to data types.

```

PurchaseDay IS NOT NULL
AND
(
    PurchaseYear IS NOT NULL
    AND
    (
        UnitCost IS NOT NULL
        AND
        (
            Quantity IS NOT NULL
            AND
            (
                Currency IS NOT NULL
                AND
                (
                    BranchPlantKey IS NOT NULL
                    AND
                    (
                        ItemMasterKey IS NOT NULL
                        AND
                        (
                            CustVendorKey IS NOT NULL
                            AND
                            (
                                TransTypeKey IS NOT NULL
                                AND
                                PurchaseMonth IS NOT NULL
                            )
                        )
                    )
                )
            )
        )
    )
)

```

Step 2: DataValidator (Check Data Type)

Data Validator

Stepname : Data Validator

Select a validation to edit :

- BranchPlantKey
- Date
- ItemMasterKey
- TransTypeKey
- CustVendorKey
- UnitCost
- Quantity
- Currency

☐ Report all errors, not only the first

☐ Output one row, concatenate errors with separator : |

Validation description BranchPlantKey

Name of field to validate BranchPlantKey

Error code

Error description

Type

Verify data type? ☒

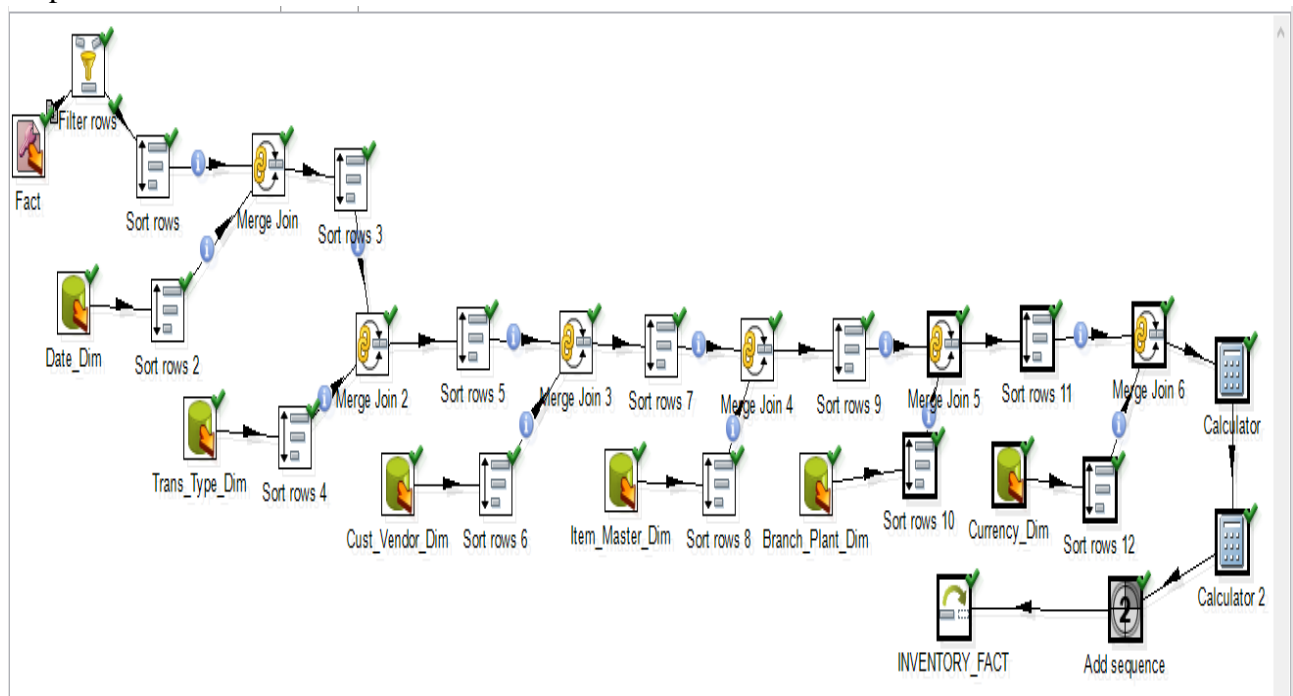
Data type Integer

Conversion mask

Decimal Symbol

Grouping Symbol

Step 3: Transformation



1. Calculator1 => represents $\text{Unit_Cost(Inserting)} = \text{Unit_Cost} * \text{Exchange_Rate}$
 2. Calculator2 => represents $\text{Ext_cost} = \text{Unit_Cost(calculated from Calculator 1)} * \text{Quantity}$
- Output: 11 rows will be affected.

Execution History Logging Step Metrics Performance Graph Metrics Preview data												
#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)
7	Sort rows 3	0	12	12	0	0	0	0	0	Finished	0.5s	24
8	Merge Join 2	0	17	12	0	0	0	0	0	Finished	0.9s	20
9	Sort rows 4	0	5	5	0	0	0	0	0	Finished	0.0s	200
10	Trans_Type_Dim	0	0	5	5	0	0	0	0	Finished	0.0s	135
11	Branch_Plant_Dim	0	0	20	20	0	0	0	0	Finished	0.0s	426
12	Sort rows 10	0	20	20	0	0	0	0	0	Finished	0.0s	488
13	Currency_Dim	0	0	4	4	0	0	0	0	Finished	0.0s	114
14	Sort rows 12	0	4	4	0	0	0	0	0	Finished	0.0s	138
15	Cust_Vendor_Dim	0	0	20	20	0	0	0	0	Finished	0.1s	370
16	Sort rows 6	0	20	20	0	0	0	0	0	Finished	0.1s	351
17	Item_Master_Dim	0	0	20	20	0	0	0	0	Finished	0.1s	392
18	Sort rows 8	0	20	20	0	0	0	0	0	Finished	0.0s	444
19	Merge Join 3	0	32	11	0	0	0	0	0	Finished	1.2s	26
20	Sort rows 7	0	11	11	0	0	0	0	0	Finished	1.3s	9
21	Merge Join 4	0	31	11	0	0	0	0	0	Finished	1.6s	19
22	Sort rows 9	0	11	11	0	0	0	0	0	Finished	1.7s	7
23	Merge Join 5	0	31	11	0	0	0	0	0	Finished	2.0s	15
24	Sort rows 11	0	11	11	0	0	0	0	0	Finished	2.1s	5
25	Merge Join 6	0	15	11	0	0	0	0	0	Finished	2.4s	6
26	Sort rows 5	0	12	12	0	0	0	0	0	Finished	0.9s	14
27	Calculator	0	11	11	0	0	0	0	0	Finished	2.4s	4
28	Calculator 2	0	11	11	0	0	0	0	0	Finished	2.4s	4
29	Add sequence	0	11	11	0	0	0	0	0	Finished	2.5s	4
30	INVENTORY_FACT	0	11	11	11	11	0	0	0	Finished	2.5s	4

Step 4: Loading ,we can see there are 11 new rows, that are get inserted in the database.

INVENTORYKEY	BRANCHPLANTKEY	DATEKEY	ITEMMASTERKEY	TRANSTYPEKEY	CUSTVENDORKEY	UNITCOST	QUANTITY	EXTCOST
18333	18	16	20	2	8	22	90	1980
18332	16	5	4	2	6	14.45	80	1156
18331	12	1	1	2	12	24.3	60	1458
18330	8	8	6	4	10	25.36	65	1648.4
18329	2	2	2	2	4	45.2	10	452
18328	18	13	12	2	2	1.826	25	45.65
18327	14	3	2	4	10	2.689	55	147.9
18326	6	7	6	4	14	24.178	30	725.34
18325	2	6	4	4	8	21.7448	75	1630.86
18324	16	11	10	2	12	29.078	45	1308.51
18323	10	14	14	4	18	76.3	50	3815